JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 49, 1–23 (1975)

Prediction and Smoothing for Partially Observed Markov Chains

MATS RUDEMO*

Research Institute of National Defense, Stockholm, Sweden

Submitted by K. J. Åström

Consider a continuous time Markov chain with stationary transition probabilities. A function of the state is observed. A regular conditional probability distribution for the trajectory of the chain, given observations up to time t, is obtained. This distribution also corresponds to a Markov chain, but the conditional chain has nonstationary transition probabilities. In particular, computation of the conditional distribution of the state at time s is discussed. For $s > t$, we have prediction (extrapolation), while $s < t$ corresponds to smoothing (interpolation). Equations for the conditional state distribution are given on matrix form and as recursive differential equations with varying s or t. These differential equations are closely related to Kolmogorov's forward and backward equations. Markov chains with one observed and one unobserved component are treated as a special case. In an example, the conditional distribution of the change-point is derived for a Poisson process with a changing intensity, given observations of the Poisson process.

1. INTRODUCTION

Let $\{\xi_t\}$ and $\{\eta_t\}$ be two stochastic processes defined on the same probability space. Suppose that we observe ξ_u, $u \leqslant t$, and that we want to estimate η_s, that is, compute the conditional distribution of η_s. For $s = t$ we have filtering, for $s > t$ prediction or extrapolation, and for $s < t$ smoothing or interpolation. For efficient computation it is desirable to have recursive equations. For the filter estimate this means that when we compute the estimate of $\eta_{t+\delta}$ based on ξ_u, $u \leqslant t + \delta$ for $\delta > 0$, we shall update the estimate of η_t based on ξ_u, $u \leqslant t$, and the required amount of computation should be small, for small δ. For prediction and smoothing, we shall in this paper derive equations that are recursive in s for fixed t, recursive in t for fixed s, and thirdly recursive in t for $s - t = \pm\varDelta$ with $\varDelta > 0$ fixed.

In [1], Wiener considered covariance stationary processes $\{\xi_t\}$ and $\{\eta_t\}$, with rational spectral densities, and observations of $\{\xi_t\}$ extended to the

* Present address: The Royal Veterinary and Agricultural University, Institute of Mathematics and Statistics, Copenhagen, Denmark.

infinite past. Recursive equations were obtained for the optimal linear least-squares filter estimate and for prediction and smoothing with fixed lead and lag time. The solutions were given on a form suitable for implementation with analog filters.

Recursive equations suitable for implementation with real-time digital computers were obtained by Kalman and Bucy in [2] and [3]. They considered optimal linear least-squares filtering and prediction for finite-dimensional time-varying systems, observed on a finite time interval. Smoothing was considered later (see for instance [4, 5 and 6]).

Optimal nonlinear estimation for Markov processes was first discussed by Stratonovich in [7] (see also [8]). A thorough treatment of recursive equations of several kinds for prediction and smoothing was given by Liptser and Shiryaev in [9, 10, and 11]. Two cases were considered, namely a pair of diffusion processes, and a pair consisting of a Markov chain and a diffusion process. The last model resembles the model discussed in the present paper and will therefore be discussed in some detail. The notation, which differs somewhat from that used in [9–11], is chosen to be similar to the notation used later in the paper.

Let $\{(\zeta_t, \xi_t)\}$ be a continuous time Markov process such that $\{\zeta_t\}$ is a Markov chain and $\{\xi_t\}$ a diffusion process, satisfying the Ito differential equation

$$d\xi_t = A(\xi_t, \zeta_t, t)\, dt + B(\xi_t, t)\, dw_t, \tag{1.1}$$

where $\{w_t\}$ is a Wiener process and A and B satisfy certain regularity conditions. In [9], recursive equations are obtained for $P(\zeta_s = k \mid \xi_u, u \leqslant t)$, $s > t$, for fixed t and increasing s, and for fixed s and increasing t. In [11], filter equations and smoothing equations for $P(\zeta_s = k \mid \xi_u, u \leqslant t)$, $s < t$, are obtained for fixed t and decreasing s and for fixed s and increasing t. Similar equations are also derived for the conditional transition probabilities $P(\zeta_s = k \mid \xi_u, u \leqslant t, \zeta_t = i)$, $s < t$.

The model discussed in this paper is the same as the model in [12], where filtering is considered. Let $\{\eta_t\}$ be a continuous time Markov chain with stationary transition probabilities. A function

$$\xi_t = g(\eta_t) \tag{1.2}$$

is observed. The model and some of the results in [12] are shortly described in Section 2.

In Section 3, a version of the conditional distribution for $\{\eta_s: 0 \leqslant s < \infty\}$ given $\{\xi_u: 0 \leqslant u \leqslant t\}$ is obtained. Let P^t denote this conditional distribution. The process $\{\eta_s\}$ is a Markov chain also with respect to P^t. For $s \geqslant t$, the transition probabilities of the conditioned chain coincide with the transition probabilities of the original chain. For $s < t$, the conditioned chain has nonstationary transition probabilities, which are best described in terms of reversed time. Note that a Markov process remains a Markov process when time is reversed [13, p. 83].

Recursive equations are given for prediction in Section 4 and for smoothing in Section 5. The important special case of a Markov chain with an unobserved component is discussed in Section 6. The smoothing equations for this model are simpler than for the general model (1.2). In Section 7, an example with a Poisson process with a changing intensity is discussed, and in Section 8 results from simulations are given. It is supposed that a change in the intensity occurs at a random time. Given observations of the events of the Poisson process, the conditional distribution of the change-point is derived. The filter estimate for this doubly stochastic Poisson process is derived in [14] and [15].

As discussed in [12] and [15], the models studied there are related to the doubly stochastic Poisson processes in Snyder [16]. Smoothing for these processes is discussed in [17], where several types of recursive smoothing equations are derived for the conditional distribution of the state of a diffusion process, given that an associated doubly stochastic Poisson process is observed.

2. Partially Observed Markov Chains

Let $\{\eta_t : t \in [0, \infty)\}$ be a continuous time Markov chain with finite or countably infinite state space S and stationary transition probabilities

$$P_{ik}(t) = P(\eta_{t+s} = k \mid \eta_s = i), \tag{2.1}$$

i, $k \in S$, $t \geqslant 0$. Further, the sample functions of $\{\eta_t\}$ are supposed to be right continuous with a finite number of jumps in every finite time interval. Then the transition probabilities are differentiable and satisfy Kolmogorov's forward and backward equations

$$P'_{ik}(t) = \sum_j P_{ij}(t)\, q_{jk}\,, \tag{2.2}$$

$$P'_{ik}(t) = \sum_j q_{ij} P_{jk}(t), \tag{2.3}$$

where

$$q_{ik} = P'_{ik}(0),$$

see [18] or [19]. The absolute probabilities

$$p_k(t) = P(\eta_t = k), \qquad k \in S,$$

satisfy the integrated version of the forward equations

$$p_k(t) - p_k(s) = \int_s^t \sum_j p_j(u)\, q_{jk}\, du, \tag{2.4}$$

symbolically written

$$dp_k(t) = \sum_j p_j(t)\, q_{jk}\, dt, \tag{2.5}$$

(see [12, Appendix]). Using vector-matrix notation, (2.5) may be written

$$dp(t) = p(t)\, Q\, dt, \tag{2.6}$$

where Q is the matrix with elements q_{ik}, and $p(t)$ is the row vector with components $p_k(t)$. In general, if x and f are vectors (or matrices), an equation of the type

$$dx(t) = f(x, t)\, dt \tag{2.7}$$

is here interpreted in the following way for each component (or matrix element).

$$x_k(t) - x_k(s) = \int_s^t f_k(x(u), u)\, du. \tag{2.8}$$

If $f_k(x(\cdot), \cdot)$ is continuous, the equation can of course be interpreted as a differential equation in the ordinary sense. We note that if x and y are real-valued, then

$$d(xy) = dx\, y + x\, dy, \tag{2.9}$$

which is a formulation of the formula for integration by parts for Lebesgue integrals (compare [12, Sect. 6]). For vectors and matrices with nonnegative components and elements, (2.9) is valid also for a vector-matrix or matrix product xy. In this form (2.9) will be used later in the paper.

Suppose we observe a function

$$\xi_t = g(\eta_t)$$

of the Markov chain $\{\eta_t\}$. In [12], equations are derived for the filter estimate $P(\eta_t = k \mid \xi_u, u \leqslant t)$, $k \in S$. Put $B = g(S)$ and $S_a = \{i \in S: g(i) = a\}$ for $a \in B$. Let $P(a, t)$, $a \in B$, and $Q(a, b)$, $a, b \in B$, $a \neq b$, be the matrices on $S \times S$ with elements

$$P_{ik}(a, t) = P(\eta_{t+s} = k, \xi_u = a, s \leqslant u \leqslant t + s \mid \eta_s = i),$$
$$Q_{ik}(a, b) = q_{ik}, \qquad (i, k) \in S_a \times S_b,$$

and $Q_{ik}(a, b) = 0$ otherwise. Further let N be an operator on row vectors p, satisfying $0 < \Sigma p_i < \infty$, defined by $(pN)_i = p_i/\Sigma p_k$, and put

$$\hat{p}(t) = p(0)\, P(\alpha_0, \tau_1)\, Q(\alpha_0, \alpha_1)\, P(\alpha_1, \tau_2 - \tau_1)\, Q(\alpha_1, \alpha_2) \cdots Q(\alpha_{N_t - 1}, \alpha_{N_t}) \times P(\alpha_{N_t}, t - \tau_{N_t})\, N, \tag{2.10}$$

where N_t is the number of jumps of $\{\xi_u : 0 < u \leqslant t\}$, $\tau_1 < \tau_2 < \cdots$ are the jump epochs, $\alpha_0 = \xi_0$, and $\alpha_\nu = \zeta_{\tau_\nu}$, $\nu \geqslant 1$. Then the components of $\hat{p}(t)$ are versions of the conditional probabilities that $\eta_t = k$, $k \in S$, given the observations, that is,

$$\hat{p}_k(t) = P(\eta_t = k \mid \xi_u, u \leqslant t), \qquad k \in S,$$

with probability one. In fact, more can be said about this. The conditions for applying the normalizing operator N in (2.10) are satisfied with probability one. By possibly redefining $\hat{p}(t)$ on a set of probability zero, or perhaps better by excluding a set of probability zero, we get $\hat{p}_k(t)$, $k \in S$, defined everywhere such that they form a probability distribution over S. With this modification, $\hat{p}(t)$ forms a regular conditional probability distribution (see [19, p. 347]), for η_t given ξ_u, $u \leqslant t$.

Let us remark that in the open interval between two successive jumps of the observed process, a component $\hat{p}_k(t)$ of $\hat{p}(t)$ in (2.10) is either identically zero or strictly positive. This follows from the fact that a transition probability $P_{ik}(t)$, see (2.1), is either zero for all $t > 0$ or strictly positive for all $t > 0$ (see [18, p. 126]).

Recursive differential equations may be derived from (2.10). In intervals between jumps of $\{\xi_t\}$ we have for $g(k) = \xi_t$

$$d\hat{p}_k(t) = \sum_i \hat{p}_i(t)\, q_{ik}\, dt + \hat{\lambda}(t)\, \hat{p}_k(t)\, dt, \tag{2.11}$$

where

$$\hat{\lambda}(t) = \sum_k \hat{p}_k(t)\, \lambda_k \tag{2.12}$$

with

$$\lambda_k = \sum_{i \notin S_a} q_{ki}, \tag{2.13}$$

for $k \in S_a$ (see [12, Theor. 4]). Using vector-matrix notation, (2.11) may be written

$$d\hat{p}(t) = \hat{p}(t)\, Q'\, dt + \hat{\lambda}(t)\, \hat{p}(t)\, dt, \tag{2.14}$$

where the matrix Q' has elements

$$Q'_{ij} = q_{ij}, \qquad g(i) = g(j), \tag{2.15}$$

and $Q'_{ij} = 0$ otherwise.

At jumps we have

$$\hat{p}(\tau) = \hat{p}(\tau - 0)\, Q(\alpha, \beta)\, N, \tag{2.16}$$

where $\alpha = \xi_{\tau-0}$ and $\beta = \xi_\tau$. Further (see [12, Theor. 5]), the function $\hat{\lambda}(t)$ is almost surely integrable over finite time intervals, and if no jumps occur in (s, t), $t > s$,

$$\exp\left[-\int_s^t \hat{\lambda}(u)\, du\right] = P(N_t - N_s = 0 \mid \xi_u, u \leqslant s). \tag{2.17}$$

Let us conclude this section by noting that for a Markov chain with time-varying continuous transition intensities $q_{ik}(t)$ and finitely many jumps in every finite time interval, the forward and backward equations for the transition probabilities

$$P_{ik}(s, t) = P(\eta_t = k \mid \eta_s = i), \qquad s \leqslant t,$$

take the form

$$d_t P_{ik}(s, t) = \sum_j P_{ij}(s, t)\, q_{jk}(t)\, dt, \tag{2.18}$$

$$\frac{d}{ds} P_{ik}(s, t) = -\sum_j q_{ij}(s)\, P_{jk}(s, t), \tag{2.19}$$

(see [20 and 21]). Further, (2.5) remains valid with q_{jk} replaced by $q_{jk}(t)$.

More general Markov chains with time-varying transition probabilities and right continuous step function realizations are studied in a recent paper [22]. The chains are characterized by the family of functions

$$H_k(t) = P(\theta_1 > t \mid \eta_0 = k), \qquad k \in S, \tag{2.20}$$

where θ_1 is the epoch of the first jump, and the matrices $\pi(t)$, $t > 0$, with elements

$$\pi_{ki}(t) = P(\eta_{\theta_1} = i \mid \eta_0 = k, \theta_1 = t), \qquad i \neq k. \tag{2.21}$$

The Kolmogorov equations hold somewhat modified. For chains with continuous transition intensities we have

$$H_k(t) = \exp\left[-\int_0^t q_k(u)\, du\right], \tag{2.22}$$

where $q_k(u) = -q_{kk}(u)$ and

$$\pi_{ki}(t) = q_{ki}(t)/q_k(t). \tag{2.23}$$

3. The Conditional Distribution of the Trajectory

Let us introduce the matrices $P(0, a, t) = P(a, t)$ and for $n \geqslant 1$

$$\begin{aligned} &P(n, a_0, ..., a_n, s_1, ..., s_n, t) \\ &\quad = P(a_0, s_1) Q(a_0, a_1) P(a_1, s_2 - s_1) \cdots Q(a_{n-1}, a_n) P(a_n, t - s_n), \end{aligned} \tag{3.1}$$

where $a_i \in B$, $i = 0, ..., n$, $0 < s_1 < \cdots < s_n \leqslant t$, and the vectors

$$p(n, a_0, ..., a_n, s_1, ..., s_n, t) = p(0) P(n, a_0, ..., a_n, s_1, ..., s_n, t), \tag{3.2}$$

$n \geqslant 0$, (compare [12, Eqs. (4.6) and (4.7)]). Then $\hat{p}(t)$ in (2.10) may be written

$$\hat{p}(t) = p(N_t, \alpha_0, ..., \alpha_{N_t}, \tau_1, ..., \tau_{N_t}, t) N. \tag{3.3}$$

We suppose that those elementary events for which (2.10) and (3.3) are not defined have been removed, compare the discussion above after (2.10).

Let us now define the stochastic matrix $R(t, s)$, $s \leqslant t$. For $\hat{p}_k(t) > 0$, the kth row of $R(t, s)$ is obtained by application of the normalizing operator N to the row vector, for which the ith component, $i \in S$, is the product of

$$p_i(N_s, \alpha_0, ..., \alpha_{N_s}, \tau_1, ..., \tau_{N_s}, s) \tag{3.4}$$

and

$$P_{ik}(N_t - N_s, \alpha_{N_s}, ..., \alpha_{N_t}, \tau_{N_s+1} - s, ..., \tau_{N_t} - s, t - s). \tag{3.5}$$

The normalization consists of division of each component with

$$p_k(N_t, \alpha_0, ..., \alpha_{N_t}, \tau_1, ..., \tau_{N_t}, t). \tag{3.6}$$

For $\hat{p}_k(t) = 0$, we put $R_{ki}(t, s) = \delta_{ki}$, the Kroeneker delta function. Then the matrices $R(t, s)$, $s \leqslant t$, satisfy Chapman–Kolmogorov's equation

$$R(t, s) = R(t, u) R(u, s), \qquad s \leqslant u \leqslant t.$$

Further we put

$$\hat{p}(s \mid t) = \hat{p}(t) P(s - t), \qquad s \geqslant t, \tag{3.7}$$

where $P(t)$ is the matrix with elements (2.1), and

$$\hat{p}(s \mid t) = \hat{p}(t)\, R(t, s), \qquad s \leqslant t. \tag{3.8}$$

The following theorem gives the conditional distribution of the trajectory of the Markov chain $\{\eta_s: s \geqslant 0\}$, given observations of $\xi_u = g(\eta_u)$, $0 \leqslant u \leqslant t$.

THEOREM 1. *Let P^t be a probability distribution for $\{\eta_s: s \geqslant 0\}$, which makes $\{\eta_s\}$ a Markov chain, with stationary transition probabilities for $s \geqslant t$ coinciding with the unconditional transition probabilities* (2.1), *with nonstationary backwards transition probabilities for $s \leqslant t$, given by*

$$P^t(\eta_s = i \mid \eta_u = k) = R_{ki}(u, s)$$

for $s \leqslant u \leqslant t$, where R is the stochastic matrix function defined above, and with distribution $\hat{p}(t)$ for η_t. Then P^t is a regular conditional probability distribution for $\{\eta_s: s \geqslant 0\}$ given $\{\xi_u: 0 \leqslant u \leqslant t\}$. In particular, $\hat{p}(s \mid t)$ defined by (3.7) *and* (3.8) *is a regular conditional probability distribution for η_s given $\{\xi_u: 0 \leqslant u \leqslant t\}$.*

Proof. The proof of Theorem 1 above is an extension of the proof of Theorem 2 in [12] and we shall start with a short description of that proof. According to Theorem 2 in [12], $\hat{p}_k(t)$ as given by (2.10) is the conditional probability that $\eta_t = k$, given $\{\xi_u: 0 \leqslant u \leqslant t\}$. This means that

$$P(\eta_t = k, A) = \int_A \hat{p}_k(t)\, dP, \tag{3.9}$$

for all $A \in \mathscr{F}_t$, the σ-algebra generated by $\{\xi_u: 0 \leqslant u \leqslant t\}$. By use of a uniqueness theorem for finite measures (see for instance [19, p. 335]), it is shown in [12] that it is enough to prove (3.9) for special sets A of the form

$$A = \{N_t = n, \xi_0 = a_0, \xi_{\tau_\nu} = a_\nu, u_\nu < \tau_\nu \leqslant v_\nu, \nu = 1,\ldots, n\}, \tag{3.10}$$

where $n \geqslant 0$, $a_\nu \in B$, $\nu = 0,\ldots, n$ and

$$0 \leqslant u_1 < v_1 \leqslant u_2 < \cdots \leqslant u_n < v_n \leqslant t. \tag{3.11}$$

The proof of (3.9) for special A is based on the following relation [12, Lemma 3]. Subject to (3.11) we have

$$\begin{aligned} &P(N_t = n, \xi_0 = a_0, \xi_{\tau_\nu} = a_\nu, u_\nu < \tau_\nu \leqslant v_\nu, \nu = 1,\ldots, n, \eta_t = k) \\ &\quad = \int_{u_1}^{v_1} \cdots \int_{u_n}^{v_n} p_k(n, a_0,\ldots, a_n, s_1,\ldots, s_n, t)\, ds_1 \cdots ds_n . \end{aligned} \tag{3.12}$$

This relation plays a dual role in the proof of (3.9). First, it gives the left member of (3.9). Second, after summation with respect to k, it gives the joint distribution of N_t, $\{\tau_\nu\}$ and $\{\xi_{\tau_\nu}\}$, and thus it may be used to evaluate the right member of (3.9), since $\hat{p}_k(t)$ is a function of these random variables.

Let us now proceed to the proof of Theorem 1. Let $\mathscr{H}$ denote the σ-algebra generated by $\{\eta_s: s \geqslant 0\}$. We have to show that for $H \in \mathscr{H}$ and $A \in \mathscr{F}_t$

$$P(H \cap A) = \int_A P^t(H)\, dP. \tag{3.13}$$

Using the aforementioned uniqueness theorem for finite measures, we get that it is enough to prove (3.13) for special sets H of the form

$$H = \{\eta_{t_0} = i_0, \ldots, \eta_{t_m} = i_m, \eta_t = k, \eta_{y_1} = j_1, \ldots, \eta_{y_r} = j_r\}, \tag{3.14}$$

where

$$0 = t_0 < \cdots < t_m < t < y_1 < \cdots < y_r,$$

and special sets A of the form (3.10) subject to (3.11) and the additional constraint that

$$t_j \notin (u_\nu, v_\nu), \qquad j = 1,\ldots, m, \quad \nu = 1,\ldots, n. \tag{3.15}$$

To prove (3.13) for special H and A, we shall use the following method. First, we derive a formula similar to (3.12) for $P(H \cap A)$. This gives the left member of (3.13). Further, $P^t(H)$ is a function of N_t, $\{\tau_\nu\}$, and $\{\xi_{\tau_\nu}\}$, and the right member of (3.13) will be evaluated by use of the joint distribution of these variables as given by (3.12). The main difficulties in the proof of (3.13) arise from the conditions on η_s for $0 \leqslant s < t$ in (3.14), while the conditions corresponding to η_s for $s > t$ are easily dealt with. To simplify notations we shall suppose that $r = 0$, that is, we consider special sets H of the form

$$H = \{\eta_{t_0} = i_0, \ldots, \eta_{t_m} = i_m, \eta_t = k\}, \tag{3.16}$$

with

$$0 = t_0 < \cdots < t_m < t. \tag{3.17}$$

Let us introduce the shorthand notation

$$\mathbf{P}_{i_j i_{j+1}} = P_{i_j i_{j+1}}(n_{j+1} - n_j, a_{n_j}, \ldots, a_{n_{j+1}}, s_{n_j+1} - t_j, \ldots, s_{n_{j+1}} - t_j, t_{j+1} - t_j),$$

where n_j is the number of jumps of $\{\xi_u: 0 < u \leqslant t_j\}$ as specified by A in (3.10) subject to (3.11) and (3.15), $j = 0,\ldots, m$, with $t_{m+1} = t$ and $i_{m+1} = k$.

Then it follows from repeated application of (3.12), and the Markov property of $\{\eta_s\}$, that for these special H and A

$$P(H \cap A) = \int_{u_1}^{v_1} \cdots \int_{u_n}^{v_n} p_{i_0}(0)\, \mathbf{P}_{i_0 i_1} \cdots \mathbf{P}_{i_m k}\, ds_1 \cdots ds_n . \tag{3.18}$$

Further,

$$P^t(H) = \hat{p}_k(t)\, R_{k i_m}(t, t_m) \cdots R_{i_1 i_0}(t_1 , t_0).$$

For $N_t = n$, $\xi_0 = a_0$, $\xi_{\tau_\nu} = a_\nu$, $\tau_\nu = s_\nu$, $\nu = 1,\dots, n$, it follows from the definition of $R_{ki}(t, s)$ (see (3.4), (3.5) and (3.6)) that on A

$$P^t(H) = \mathbf{P}_{i_m k} \cdots \mathbf{P}_{i_0 i_1} p_{i_0}(0) \Big/ \sum_i p_i(n, a_0 ,\dots, a_n , s_1 ,\dots, s_n , t).$$

By use of (3.12) we find that

$$\int_A P^t(H)\, dP = \int_{u_1}^{v_1} \cdots \int_{u_n}^{v_n} \mathbf{P}_{i_m k} \cdots \mathbf{P}_{i_1 i_0} p_{i_0}(0)\, ds_1 \cdots ds_n ,$$

which coincides with the right member of (3.18). This gives (3.13) and thus Theorem 1 is proved. Q.E.D.

Remark 1. In [12, Sect. 2], filtering for discrete time Markov chains was briefly discussed. In particular it was shown that the row vector

$$\hat{p}(t) = p(0)\, P(\xi_0 , \xi_1)\, P(\xi_1 , \xi_2) \cdots P(\xi_{t-1} , \xi_t)\, N \tag{3.19}$$

gives a regular conditional probability distribution for η_t, given ξ_u, $u \leqslant t$. Here $P(a, b)$, a, $b \in g(S)$, is the substochastic matrix obtained from the original transition probability matrix P for $\{\eta_t\}$, $P_{ik} = P\{\eta_{t+1} = k \mid \eta_t = i\}$, by replacing all matrix elements with indices $(i, k) \notin S_a \times S_b$ with zeros. As in the continuous time case, a function $\xi_t = g(\eta_t)$, $t = 0, 1,\dots$, is supposed to be observed. In this remark we shall give without proof the result corresponding to Theorem 1 for discrete time.

Let $\xi_0 , \xi_1 ,\dots, \xi_t$ be given, and let $R(s, s-1)$, $1 \leqslant s \leqslant t$, be a stochastic matrix with elements

$$R_{ki}(s, s-1) = \frac{\hat{p}_i(s-1)\, P_{ik}}{\sum_j \hat{p}_j(s-1)\, P_{jk}} , \tag{3.20}$$

for all k such that $\hat{p}_k(s) > 0$. Further, let P^t be the probability measure which makes $\{\eta_s : s \geqslant 0\}$ a Markov chain with stationary transition probability matrix P for $s \geqslant t$, with nonstationary backwards transition probabilities

$$P^t(\eta_{s-1} = i \mid \eta_s = k) = R_{ki}(s, s-1) \tag{3.21}$$

for $s \leqslant t$, and distribution $\hat{p}(t)$ for η_t. Then P^t is a regular conditional probability distribution for $\{\eta_s: s \geqslant 0\}$ given $\{\xi_u: 0 \leqslant u \leqslant t\}$. In particular $\hat{p}(s \mid t)$ defined by

$$\hat{p}(s \mid t) = \hat{p}(t) P^{s-t}, \qquad s \geqslant t, \tag{3.22}$$

and

$$\hat{p}(s \mid t) = \hat{p}(t) R(t, t-1) \cdots R(s+1, s), \qquad s \leqslant t, \tag{3.23}$$

is a regular conditional probability distribution for η_s given $\{\xi_u: 0 \leqslant u \leqslant t\}$.

4. Recursive Equations for Prediction

According to Theorem 1,

$$\hat{p}(s \mid t) = \hat{p}(t) P(s-t), \qquad s \geqslant t, \tag{4.1}$$

gives the conditional distribution of the state of the Markov chain at a future time point s, given observations up to time t. In this section we shall derive recursive equations in s and t for $\hat{p}(s \mid t)$ in (4.1). Let Q'' be the matrix with elements

$$Q''_{ij} = q_{ij}, \qquad g(i) \neq g(j), \tag{4.2}$$

and $Q''_{ij} = 0$ if $g(i) = g(j)$. Note that

$$Q = Q' + Q'', \tag{4.3}$$

where Q' is the matrix defined by (2.15). Then the following theorem is valid.

Theorem 2. *For fixed t,*

$$d_s \hat{p}(s \mid t) = \hat{p}(s \mid t) Q \, ds. \tag{4.4}$$

Further for fixed s,

$$d_t \hat{p}(s \mid t) = \hat{\lambda}(t) \hat{p}(s \mid t) \, dt - \hat{p}(t) Q'' P(s-t) \, dt \tag{4.5}$$

in intervals between jumps. For fixed $\Delta > 0$,

$$d_t \hat{p}(t+\Delta \mid t) = \hat{\lambda}(t) \hat{p}(t+\Delta \mid t) \, dt + \hat{p}(t) Q' P(\Delta) \, dt \tag{4.6}$$

in intervals between jumps.

Remark 2. Suppose we have observed ξ_u, $u \leqslant t$, and that we want to estimate η_s for all $s \geqslant t$. Then we can use (4.4) for recursive computation with the initial value $\hat{p}(s \mid t) = \hat{p}(t)$ for $s = t$. Suppose on the other hand that we are especially interested in η_s for a special value of s. To see how the estimate $\hat{p}(s \mid t)$ improves as the observation time t increases towards s, we can use (4.5) in intervals between jumps. After a jump epoch τ we restart with the initial value $\hat{p}(s \mid t) = \hat{p}(\tau)\, P(s - \tau)$ for $t = \tau$. It should however be noted that a direct evaluation of $\hat{p}(s \mid t)$ from (4.1) often should be preferable to the above described use of (4.5). Similar remarks are valid for the fixed-lead prediction via (4.6) or alternatively from $\hat{p}(t + \Delta \mid t) = \hat{p}(t)\, P(\Delta)$.

Proof of Theorem 2. Equation (4.4) follows from Kolmogorov's forward equation (2.6). To derive (4.5), we shall use (4.1), (2.9), (2.14) and the integrated form of Kolmogorov's backward equation (2.3) for $P(s - t)$, which may be written

$$d_t P(s - t) = -QP(s - t)\, dt.$$

We get

$$\begin{aligned} d_t \hat{p}(s \mid t) &= d\hat{p}(t)\, P(s - t) + \hat{p}(t)\, d_t P(s - t) \\ &= \hat{\lambda}(t)\, \hat{p}(t)\, P(s - t)\, dt + \hat{p}(t)\, Q' P(s - t)\, dt - \hat{p}(t)\, QP(s - t)\, dt \\ &= \hat{\lambda}(t)\, \hat{p}(s \mid t)\, dt - \hat{p}(t)\, Q'' P(s - t)\, dt. \end{aligned}$$

In the last step we have used (4.3). Finally (4.6) is obtained by addition of the right members of (4.4) and (4.5) with $s = t + \Delta$, and with Q'' in (4.5) written as $Q - Q'$. Note that $P(\Delta)$ and Q commute. This concludes the proof of Theorem 2. Q.E.D.

5. Recursive Equations for Smoothing

From Theorem 1, we find that

$$\hat{p}(s \mid t) = \hat{p}(t)\, R(t, s), \qquad s \leqslant t, \tag{5.1}$$

gives the conditional distribution of the state of the Markov chain at a previous time point s, given observations up to time t. We shall now derive recursive equations in s and t for $\hat{p}(s \mid t)$ in (5.1) and for $R(t, s)$.

For $\hat{p}_k(s) > 0$, the elements in the kth row of the matrix $Q^R(s)$ on $S \times S$ are defined by

$$q^R_{ki}(s) = q_{ik}\hat{p}_i(s)/\hat{p}_k(s), \qquad i \neq k, \tag{5.2}$$

and

$$q_{kk}^R(s) = -\frac{1}{\hat{p}_k(s)} \sum_{i \neq k} \hat{p}_i(s)\, q_{ik} \,. \tag{5.3}$$

For $\hat{p}_k(s) = 0$, we put $q_{ki}^R(s) = 0$. Further for a jump epoch τ of the observed process, we let $J(\tau)$ be the matrix on $S \times S$ with elements

$$J_{ki}(\tau) = \hat{p}_i(\tau - 0)\, q_{ik} \Big/ \sum_j \hat{p}_j(\tau - 0)\, q_{jk} \,, \tag{5.4}$$

when $\hat{p}_k(\tau) > 0$ and $J_{ki}(\tau) = \delta_{ki}$ otherwise. The following theorem gives recursive equations in s.

THEOREM 3. *Let t be fixed. In an s-interval with $s \leqslant t$ and without jumps of the observed process, we have*

$$d_s R(t, s) = -R(t, s)\, Q^R(s)\, ds \tag{5.5}$$

and

$$d_s \hat{p}(s \mid t) = -\hat{p}(s \mid t)\, Q^R(s)\, ds. \tag{5.6}$$

At a jump epoch $\tau \leqslant t$, we have

$$R(t, \tau - 0) = R(t, \tau)\, J(\tau) \tag{5.7}$$

and

$$\hat{p}(\tau - 0 \mid t) = \hat{p}(\tau \mid t)\, J(\tau). \tag{5.8}$$

Proof. Let us use $\mathbf{p}_i(s)$, $\mathbf{P}_{ik}(s, t)$ and $\mathbf{p}_k(t)$ as shorthand notations for the variables in (3.4), (3.5) and (3.6). Then, for $\hat{p}_k(t) > 0$,

$$R_{ki}(t, s) = \mathbf{p}_i(s)\, \mathbf{P}_{ik}(s, t) / \mathbf{p}_k(t). \tag{5.9}$$

Further, for $\hat{p}_i(s) > 0$, we have $\hat{p}_j(s)/\hat{p}_i(s) = \mathbf{p}_j(s)/\mathbf{p}_i(s)$ and (compare (2.5))

$$d_s \mathbf{p}_i(s) = \sum_j \mathbf{p}_j(s)\, q_{ji}\, ds = \mathbf{p}_i(s)\, [-q_{ii}^R(s) + q_{ii}]\, ds. \tag{5.10}$$

Similarly we get (compare (2.3))

$$\begin{aligned} d_s \mathbf{P}_{ik}(s, t) &= -\sum_j q_{ij} \mathbf{P}_{jk}(s, t) \\ &= -q_{ii} \mathbf{P}_{ik}(s, t) - \sum_{j \neq i} q_{ji}^R(s)\, \mathbf{p}_j(s)\, \mathbf{P}_{jk}(s, t) / \mathbf{p}_i(s). \end{aligned} \tag{5.11}$$

It follows from (5.9) that

$$d_s R_{ki}(t, s) = [d_s \mathbf{p}_i(s) \mathbf{P}_{ik}(s, t) + \mathbf{p}_i(s)\, d_s \mathbf{P}_{ik}(s, t)]/\mathbf{p}_k(t)$$
$$= -\sum_j R_{kj}(t, s)\, q_{ji}^R(s),$$

which gives (5.5). Multiplication of (5.5) from the left with $\hat{p}(t)$ gives (5.6).

To prove (5.7), we note that $J_{ki}(\tau)$ in (5.4) may be written

$$J_{ki}(\tau) = \mathbf{p}_i(\tau - 0)\, q_{ik}/\mathbf{p}_k(\tau).$$

By use of this equation, we get

$$R_{ki}(t, \tau - 0) = \mathbf{p}_i(\tau - 0) \mathbf{P}_{ik}(\tau - 0, t)/\mathbf{p}_k(t)$$
$$= \sum_j \mathbf{p}_j(\tau)\, J_{ji}(\tau) \mathbf{P}_{jk}(\tau, t)/\mathbf{p}_k(t) = \sum_j R_{kj}(t, \tau)\, J_{ji}(\tau),$$

which gives (5.7). Multiplication of (5.7) from the left with $\hat{p}(t)$ gives (5.8), and Theorem 3 is proved. Q.E.D.

The following theorem gives recursive equations in t for fixed s and for fixed lag.

THEOREM 4. *For fixed s, let us regard a t-interval with $t \geqslant s$ and without jumps of the observed process. Then,*

$$d_s R(t, s) = Q^R(t)\, R(t, s)\, dt \tag{5.12}$$

and

$$d_t \hat{p}(s \mid t) = \hat{\lambda}(t)\, \hat{p}(s \mid t)\, dt - \hat{p}(t)\, \Lambda R(t, s)\, dt, \tag{5.13}$$

where Λ is the diagonal matrix with diagonal elements λ_k, $k \in S$ (see (2.13)). *Further, for fixed $\Delta > 0$, let us regard a t-interval (t_1, t_2) such that neither (t_1, t_2), nor $(t_1 - \Delta, t_2 - \Delta)$, contains a jump of the observed process. In this interval,*

$$\begin{aligned} d_t \hat{p}(t - \Delta \mid t) = -\hat{p}(t - \Delta \mid t)\, Q^R(t - \Delta)\, dt + \hat{\lambda}(t)\, \hat{p}(t - \Delta \mid t)\, dt \\ -\hat{p}(t)\, \Lambda R(t, t - \Delta)\, dt. \end{aligned} \tag{5.14}$$

Proof. The notations introduced in the proof of Theorem 3 will be used. In a t-interval without jumps of the observed process, we have for $\hat{p}_k(t) > 0$ (compare (2.2))

$$d_t \mathbf{P}_{ik}(s, t) = \sum_j \mathbf{P}_{ij}(s, t)\, q_{jk}\, dt. \tag{5.15}$$

Together with (5.9) and (5.10) with i and s replaced by k and t, this gives

$$\begin{aligned} d_t R_{ki}(t, s) &= \mathbf{p}_i(s)\, d_t \mathbf{P}_{ik}(s, t)/\mathbf{p}_k(t) - \mathbf{p}_i(s)\, \mathbf{P}_{ik}(s, t)\, d_t \mathbf{p}_k(t)/[\mathbf{p}_k(t)]^2 \\ &= \Big\{\sum_j \mathbf{p}_i(s)\, \mathbf{P}_{ij}(s, t)\, q_{jk}/\mathbf{p}_k(t) - R_{ki}(t, s)\, [-q^R_{kk}(t) + q_{kk}]\Big\}\, dt \\ &= \sum_j q^R_{kj}(t)\, R_{ji}(t, s), \end{aligned}$$

and (5.12) follows. From (2.14) and (5.12), we find that

$$\begin{aligned} d_t \hat{p}(s \mid t) &= d\hat{p}(t)\, R(t, s)\, dt + \hat{p}(t)\, d_t R(t, s) \\ &= \hat{p}(t)\, Q' R(t, s)\, dt + \hat{\lambda}(t)\, \hat{p}(t)\, R(t, s)\, dt + \hat{p}(t)\, Q^R(t)\, R(t, s)\, dt. \end{aligned}$$

Using (5.2) and (5.3), we get that, for $\xi_t = a$ and $k \in S_a$, the kth component of $\hat{p}(t)\, Q' + \hat{p}(t)\, Q^R(t)$ is

$$\sum_{i \in S_a} \hat{p}_i(t)\, q_{ik} + \sum_{j \in S_a} \hat{p}_j(t)\, q^R_{jk}(t) = -\lambda_k \hat{p}_k(t),$$

which gives (5.13). Finally, (5.14) is obtained from (5.6) and (5.13).

Q.E.D.

Remark 3. Suppose that we have observed ξ_u for $u \leqslant t$, and that we wish to compute $\hat{p}(s \mid t)$ for all $s \leqslant t$. This is sometimes called "fixed interval smoothing" (see for instance [5]). Suppose we have computed $\hat{p}(t)$. We could then use (5.1) with $R(t, s)$ computed according to (5.9), or perhaps (5.5) and (5.7). However, with regard to the amount of computation it is usually more convenient to use (5.6) in intervals between jumps of $\{\xi_s\}$ and (5.8) at jumps. Suppose on the other hand that we are most interested in $\eta(s)$ for a particular value of s. This corresponds to "fixed point smoothing" in [5]. Then (5.13) can be used to compute the changes of $\hat{p}(s \mid t)$ in t-intervals without jumps. Further, (5.14) corresponds to "fixed lag smoothing" in [5]. However, (5.13) and (5.14) contain the matrices $R(t, s)$ and $R(t, t - \Delta)$ and are hence not as effective as (5.6) from a computation point of view. Presumably, the most effective equations for "fixed point smoothing" and "fixed lag smoothing" are obtained from computation of $R(t, s)$ via (5.12), or from recursive equations for $\mathbf{p}_i(s)$, $\mathbf{P}_{ik}(s, t)$, and $\mathbf{p}_k(t)$ in (5.9). Such equations may easily be deduced from their definitions, see (3.1)–(3.6). In fact, in (5.10), (5.11) and (5.15), such equations are given for intervals between jumps.

Remark 4. Equations (5.5) and (5.12) have been deduced from the defining equation (5.9) for $R(s, t)$. Another way to derive these equations is to use the fact that $\{R(u, s) \colon s \leqslant u \leqslant t\}$ are the nonstationary transition probability matrices for the Markov chain $\{\eta_s : t \geqslant s \geqslant 0\}$ with respect to the measure P^t. One shows that $\{q_{ki}^R(s)\}$ are the transition intensities of this chain. For instance, if no jumps of the observed process occur in $[s - h, s]$ if $\hat{p}_k(s) > 0$, if $i \neq k$, and $g(i) = g(k) = a$, one can show that

$$\begin{aligned} P^t(\eta_{s-h} = i \mid \eta_s = k) &= P^s(\eta_{s-h} = i \mid \eta_s = k) = P^s(\eta_{s-h} = i, \eta_s = k)/\hat{p}_k(s) \\ &= P^{s-h}(\eta_{s-h} = i, \eta_s = k \mid N_s = N_{s-h})/\hat{p}_k(s) \\ &= \hat{p}_i(s - h)\, P_{ik}(a, h)/[\hat{p}_k(s)\, P^{s-h}(N_s = N_{s-h})]. \end{aligned}$$

From this relation, it follows that

$$\lim_{h \to 0+} P^t(\eta_{s-h} = i \mid \eta_s = k)/h = \hat{p}_i(s)\, q_{ik}/\hat{p}_k(s) = q_{ki}^R(s). \tag{5.16}$$

Equations (5.5) and (5.12) are the forward and backward Kolmogorov equations for a chain with nonstationary transition intensities $\{q_{ki}^R(s)\}$ (compare the discussion in the end of Section 2). It may be noted that (5.5) and (5.12) are very similar to equations in [11] for the transition probabilities $P(\zeta_s = k \mid \xi_u, u \leqslant t, \zeta_t = i)$, $s < t$, for the model (1.1), though the derivations are quite different.

Let us also remark that between jumps of the observed process, the conditional transition intensities in the forward direction are given by

$$q_{ik}^F(s) = \lim_{h \to 0+} P^t(\eta_{s+h} = k \mid \eta_s = i)/h = \frac{\hat{p}_k(s \mid t)}{\hat{p}_k(s)} \frac{\hat{p}_i(s)}{\hat{p}_i(s \mid t)}\, q_{ik}, \tag{5.17}$$

for $s < t$, $i \neq k$ and $\hat{p}_k(s)\, \hat{p}_i(s \mid t) > 0$. Note that $q_{ik}^F(s)$ depends on t while $q_{ki}^R(s)$ is independent of t, and that $q_{ik}^F(s) \to q_{ik}$ as $s \to t - 0$, provided that $\hat{p}_k(t)\, \hat{p}_i(t) > 0$.

6. Markov Chains with an Unobserved Component

Following [12, Sect. 7], we suppose here that the Markov chain $\{\eta_t\}$ has the form

$$\eta_t = (\zeta_t, \xi_t), \tag{6.1}$$

where $\{\zeta_t\}$ is an unobserved component with state space S_1, and ξ_t is the observed component with state space S_2. The probability of a simultaneous change of $\{\zeta_t\}$ and $\{\xi_t\}$ is supposed to be zero. As in [12, Sect. 7], we let $\{q_{ki}(a) : (k, i) \in S_1 \times S_1, a \in S_2\}$ and $\{\lambda_{ab}(k) : (a, b) \in S_2 \times S_2, k \in S_1\}$ denote the transition intensities of the components; the matrices $P(a, t)$ and $Q(a, b)$ on $S_1 \times S_1$ are defined by [12, Eqs. (7.1) and (7.2)]. Then the row vector

$\hat{p}(t)$ on S_1, defined by (2.10) with $p(0)$ as the initial distribution of ζ_0 over S_1, becomes a regular conditional probability distribution for ζ_t, given $\{\xi_u: 0 \leqslant u \leqslant t\}$. Similarly $R(t, s)$ (see (3.1)–(3.6) and (5.9)) becomes a matrix on $S_1 \times S_1$. The restriction of the measure P^t in Theorem 1 to $\{\zeta_s: 0 \leqslant s \leqslant t\}$ specifies a Markov chain with nonstationary transition probabilities corresponding to the distribution $\hat{p}(t)$ for ζ_t and the transition probabilities (in reversed time) given by $R(u, s)$, $0 \leqslant s \leqslant u \leqslant t$. If $\{q_{ki}(a)\}$ are independent of a, $\{\zeta_s: 0 \leqslant s < \infty\}$ forms a Markov chain with respect to P. Then $\{\zeta_s: t \leqslant s < \infty\}$ also forms a Markov chain with respect to P^t with the same transition probabilities.

The most important simplification caused by consideration of processes of the type (6.1) concerns smoothing. If $\hat{p}(t)$ on $R(t, s)$, $s \leqslant t$ are defined as above, then

$$\hat{p}(s \mid t) = \hat{p}(t)\, R(t, s)$$

becomes the conditional distribution for ζ_s given $\{\xi_u: 0 \leqslant u \leqslant t\}$. Let $Q^R(s)$ be defined by (5.2) and (5.3) with $(i, k) \in S_1 \times S_1$ and q_{ik} replaced by $q_{ik}(\xi_s)$. Then (5.5) and (5.6) are valid in the whole interval $0 \leqslant s \leqslant t$, that is, $J(\tau)$ in (5.7) and (5.8) is replaced by the identity matrix. In particular,

$$\hat{p}(s \mid t) = \hat{p}(t) + \int_s^t \hat{p}(u \mid t)\, Q^R(u)\, du, \tag{6.2}$$

for $0 \leqslant s \leqslant t$. Note that the continuity of $\hat{p}(s \mid t)$ with respect to s also follows directly from the assumption that $\{\zeta_t\}$ and $\{\xi_t\}$ with probability one have no common discontinuity points.

7. A Poisson Process with Randomly Changing Intensity

Let $\{N_t: t \in [0, \infty)\}$ be a point process, such that $\{N_t\}$ for $t < \theta$ is a Poisson process with intensity λ_0, and $\{N_t\}$ for $t > \theta$ is a Poisson process with intensity λ_1. If θ is a stochastic variable, then $\{N_t\}$ is a doubly stochastic Poisson process. The case where θ is an exponentially distributed variable is discussed in [14] and [15]. In [23], the estimation of θ is considered with θ as a nonstochastic parameter.

Let $\{\zeta_t\}$ be a Markov chain with finite state space S and one absorbing state $1 \in S$. If q_{ki} and q_k, $k, i \in S$, are the transition intensities, then $q_1 = 0$. Let θ be the time to absorption, let $\{N_t\}$ be the doubly stochastic Poisson process described above, that is, the intensity at time t is λ_0 if $t < \theta$ and λ_1 if $t > \theta$, and let $\tau_1 < \tau_2 \cdots$ denote the epochs of increase of N_t, $t > 0$. Let further F be the distribution of the time to absorption, that is,

$$F(t) = P(\theta \leqslant t) = P(\zeta_t = 1). \tag{7.1}$$

As discussed in [12, Sect. 8] an arbitrary distribution may be approximated in this way by a suitable choice of the Markov chain.

Suppose we have observed N_u, $u \leqslant t$. Let $\hat{F}(\cdot \mid t)$ be the conditional distribution of θ, that is,

$$\hat{F}(s \mid t) = P(\theta \leqslant s \mid N_u, u \leqslant t). \tag{7.2}$$

Let $\hat{p}(s \mid t)$ be a version of the row vector of conditional probabilities for ζ_s, given N_u, $u \leqslant t$. Then,

$$\hat{F}(s \mid t) = \hat{p}_1(s \mid t). \tag{7.3}$$

To compute $\hat{F}(s \mid t)$ we can first determine $\hat{p}(t)$ by solving the system of differential equation (see [12, Eqs. (1.5)–(1.7)])

$$\hat{p}_k{}'(u) = \sum_i \hat{p}_i(u)\, q_{ik} + (\lambda_1 - \lambda_0)\, \hat{p}_1(u)\, \hat{p}_k(u), \qquad k \neq 1, \tag{7.4}$$

$$\hat{p}_1{}'(u) = \sum_i \hat{p}_i(u)\, q_{i1} - (\lambda_1 - \lambda_0)\, \hat{p}_1(u)\, [1 - \hat{p}_1(u)], \tag{7.5}$$

for $0 \leqslant u < \tau_1$, $\tau_\nu \leqslant u < \tau_{\nu+1}$, $\nu = 1,..., N_t - 1$, and $\tau_{N_t} \leqslant u \leqslant t$, with the initial values

$$\hat{p}_k(0) = P(\zeta_0 = k), \tag{7.6}$$

and for $\nu \geqslant 1$,

$$\hat{p}_k(\tau_\nu) = \hat{p}_k(\tau_\nu - 0)\, \lambda_0/\hat{\lambda}(\tau_\nu - 0), \qquad k \neq 1, \tag{7.7}$$

$$\hat{p}_1(\tau_\nu) = \hat{p}_1(\tau_\nu - 0)\, \lambda_1/\hat{\lambda}(\tau_\nu - 0), \tag{7.8}$$

with

$$\hat{\lambda}(u) = [1 - \hat{p}_1(u)]\, \lambda_0 + \hat{p}_1(u)\, \lambda_1 . \tag{7.9}$$

With $\hat{p}(t)$ thus determined, we get $\hat{p}(s \mid t)$ for $s \geqslant t$ from

$$\frac{\partial}{\partial s} \hat{p}_k(s \mid t) = \sum_i \hat{p}_i(s \mid t)\, q_{ik}, \qquad s \geqslant t, \tag{7.10}$$

with initial value $\hat{p}(t \mid t) = \hat{p}(t)$. Further with

$$q_{ki}^R(s) = q_{ik} \hat{p}_i(s)/\hat{p}_k(s), \qquad i \neq k, \tag{7.11}$$

$$q_{kk}^R(s) = -q_k{}^R(s) = - \sum_{i \neq k} q_{ki}^R(s), \tag{7.12}$$

we get $\hat{p}(s \mid t)$ for $s \leqslant t$ from

$$\frac{\partial}{\partial s} \hat{p}_k(s \mid t) = - \sum_i \hat{p}_i(s \mid t)\, q_{ik}^R(s), \qquad s \leqslant t, \tag{7.13}$$

with decreasing s and initial value $\hat{p}(t \mid t) = \hat{p}(t)$.

Let us now specialize to an exponential distribution for θ. Choose $\{\zeta_t\}$ to have two states, 0 and 1, with 1 absorbing and the transition intensity $q_{01} = q$. Further we start with $\zeta_0 = 0$. Then the distribution function of θ is

$$F(t) = P(\theta \leqslant t) = P(\zeta_t = 1) = 1 - e^{-qt}.$$

Given observations of N_u, $u \leqslant t$, we compute $\hat{p}_1(t)$ from

$$\hat{p}_1'(u) = [1 - \hat{p}_1(u)]\, [q - (\lambda_1 - \lambda_0)\, \hat{p}_1(u)],$$

with initial values $\hat{p}_1(0) = 0$ and $\hat{p}_1(\tau_\nu)$, $\nu = 1,..., N_t$, given by (7.8). Equation (7.10) for prediction becomes

$$(\partial/\partial s)\, \hat{p}_1(s \mid t) = [1 - \hat{p}_1(s \mid t)]\, q,$$

with the solution

$$\hat{p}_1(s \mid t) = 1 - [1 - \hat{p}_1(t)]\, e^{-q(s-t)}, \qquad s \geqslant t. \tag{7.14}$$

Further, from (7.11) and (7.12), we get

$$q_1{}^R(s) = q[1 - \hat{p}_1(s)]/\hat{p}_1(s).$$

Hence for $s \leqslant t$, we have (see (7.13))

$$(\partial/\partial s)\, \hat{p}_1(s \mid t) = \hat{p}_1(s \mid t)\, q_1{}^R(s),$$

with the solution

$$\hat{p}_1(s \mid t) = \hat{p}_1(t) \exp\left[- \int_s^t q_1{}^R(u)\, du\right], \qquad 0 < s \leqslant t. \tag{7.15}$$

Together with the formulas for $\hat{p}_1(t)$, Eqs. (7.14) and (7.15) determine the conditional distribution of θ, given N_u, $u \leqslant t$.

8. Results From Simulations

For the doubly stochastic Poisson process with exponential distribution for θ, described in the end of the previous section, simulations were performed. The parameters used were $q = 0.3$, $\lambda_0 = 0.5$ and $\lambda_1 = 1.5$ (compare

[15, Sect. 6]). The curves obtained for one simulated series of events are shown in Fig. 1.

Let $\hat{\theta}(t)$ and $\hat{V}(t)$ be the conditional expectation and variance of θ, given N_s, $s \leqslant t$, that is,

$$\hat{\theta}(t) = E[\theta \mid N_s, s \leqslant t], \qquad \hat{V}(t) = E[(\theta - \hat{\theta}(t))^2 \mid N_s, s \leqslant t].$$

Then

$$E[\hat{\theta}(t)] = E[\theta] = 1/q, \qquad \text{and} \qquad E[\hat{V}(t)] = \sigma^2(t),$$

where

$$\sigma^2(t) = E[(\theta - \hat{\theta}(t))^2],$$

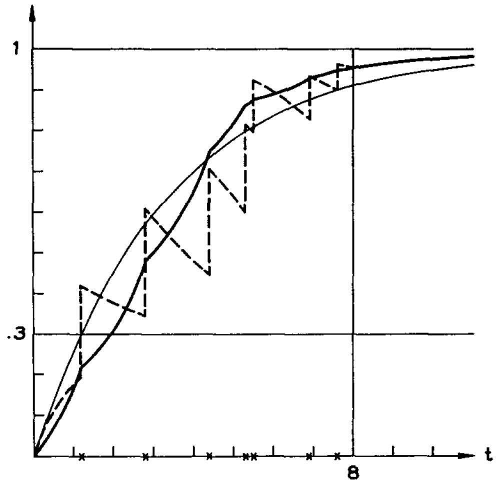

FIG. 1. The conditional distribution function $\hat{p}_1(s \mid t)$ and the filter estimate $\hat{p}_1(s)$ for a simulated series of events. The parameters are $\lambda_0 = 0.5$, $\lambda_1 = 1.5$, and $q = 0.3$, and the observation time is $t = 8$. Crosses indicate events and a transition to state 1 is supposed to have occurred at $\theta = 3$. The unconditional distribution function $1 - e^{-qs}$ is also shown. The estimates (8.1) and (8.2) assume the values $\hat{\theta}(t) = 3.40$ and $\hat{V}(t) = 6.84$.

that is, the expectation of the squared error in the least-squares estimate of θ based on the observations up to time t. As $\hat{p}_1(s \mid t)$, $0 \leqslant s < \infty$, is the conditional distribution function of θ, we have

$$\hat{\theta}(t) = \int_0^\infty [1 - \hat{p}_1(s \mid t)]\, ds \tag{8.1}$$

and

$$\hat{V}(t) = 2 \int_0^\infty s[1 - \hat{p}_1(s \mid t)]\, ds - [\hat{\theta}(t)]^2. \tag{8.2}$$

A series of $n = 10\,000$ runs was performed. Let θ_i be the θ-value, and let $\hat{\theta}_i(t)$ and $\hat{V}_i(t)$ be determined from (8.1) and (8.2) for the ith run. The variables

$$s_1^2 = \frac{1}{n} \sum_{i=1}^{n} [\hat{\theta}_i(t) - \theta_i]^2 \tag{8.3}$$

and

$$s_2^2 = \frac{1}{n} \sum_{i=1}^{n} \hat{V}_i(t) \tag{8.4}$$

were computed. Note that both s_1^2 and s_2^2 are unbiased estimates of $\sigma^2(t)$. For $t = 8$, the following values were obtained

$$s_1^2 = 5.66 \pm 0.21,$$
$$s_2^2 = 5.86 \pm 0.06,$$

where the values after the $\pm$ signs denote the standard deviations estimated from the sample. It is seen that the standard deviation of the estimate s_2^2 is about 30% of the standard deviation of the estimate s_1^2.

In Tables I and II, results obtained from variations in the observation time t and the ratio λ_1/λ_0 are shown. Only the estimates of $\sigma^2(t)$ based on $\hat{V}(t)$, that is, s_2^2, are given. For the estimates of $\sigma^2(t)$, the standard deviations computed from the samples are also given.

TABLE I

Results from Simulations with Varying Observation Time. (The parameters are $\lambda_0 = 0.5$, $\lambda_1 = 1.5$ and $q = 0.3$, and the number of runs is $n = 2500$ for each entry. The values after the $\pm$ signs denote standard deviations of the estimates.)

t	0	1	2	4	8	16
s_2^2	11.11	10.88 ± 0.02	10.13 ± 0.05	8.22 ± 0.09	5.91 ± 0.12	4.65 ± 0.12

TABLE II

Results from Simulations with Varying Intensity Ratio λ_1/λ_0. (The fixed parameters are $\lambda_0 = 0.5$, $q = 0.3$ and $t = 8$. The number of runs is $n = 2500$ for each entry. The values after the $\pm$ signs denote standard deviations of the estimates.)

λ_1/λ_0	1	2	3	5	10
s_2^2	11.11	8.90 ± 0.11	5.91 ± 0.12	2.79 ± 0.09	1.26 ± 0.06

Acknowledgments

For valuable comments, I wish to thank Karl Johan Åström, Jan Grandell and Staffan Norlander.

References

1. N. Wiener, "Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications," Wiley, New York, 1949.
2. R. E. Kalman, A new approach to linear filtering and prediction problems, *Trans. ASME Ser. D: J. Basic Eng.* **82** (1960), 34–45.
3. R. E. Kalman and R. S. Bucy, New results in linear filtering and prediction theory, *Trans. ASME Ser. D: J. Basic Eng.* **83** (1961), 95–107.
4. J. S. Meditch, On optimal linear smoothing theory, *Information and Control* **10** (1967), 598–615.
5. T. Kailath and P. Frost, An innovations approach to least-squares estimation. II. Linear smoothing in additive noise, *IEEE Trans. Automatic Control* **13** (1968), 655–660.
6. L. E. Zachrisson, On optimal smoothing of continuous time Kalman processes, *Information Sci.* **1** (1969), 143–172.
7. R. L. Stratonovich, Conditional Markov processes, *Theor. Probability Appl.* **5** (1960), 156–178.
8. R. L. Stratonovich, "Conditional Markov Processes and Their Application to the Theory of Optimal Control," Elsevier, New York, 1968.
9. R. S. Liptser and A. N. Shiryaev, The extrapolation of multidimensional Markov processes from incomplete data, *Theor. Probability Appl.* **13** (1968), 15–38.
10. R. S. Liptser and A. N. Shiryaev, Nonlinear interpolation of components of Markov diffusion processes (direct equations, effective formulas), *Theor. Probability Appl.* **13** (1968), 564–583.
11. R. S. Liptser and A. N. Shiryaev, Interpolation and filtering of a jump-like component of a Markov process, *Math. USSR-Izv.* **3** (1969), 853–865.
12. M. Rudemo, State estimation for partially observed Markov chains, *J. Math. Anal. Appl.* **44** (1973), 581–611.
13. J. L. Doob, "Stochastic Processes," Wiley, New York, 1953.
14. L. I. Galchuk and B. L. Rozovski, The "disorder" problem for a Poisson process, *Theor. Probability Appl.* **16** (1971), 712–717.
15. M. Rudemo, Doubly stochastic Poisson processes and process control, *Advances in Appl. Probability* **4** (1972), 318–338.
16. D. L. Snyder, Filtering and detection for doubly stochastic Poisson processes, *IEEE Trans. Information Theory* **18** (1972), 91–102.
17. D. L. Snyder, Smoothing for doubly stochastic Poisson processes, *IEEE Trans. Information Theory* **18** (1972), 558–562.
18. K. L. Chung, "Markov Chains with Stationary Transition Probabilities," 2nd ed., Springer, Berlin, 1967.
19. D. Freedman, "Markov Chains," Holden–Day, San Fransisco, 1971.
20. W. Feller, On the integrodifferential equations of purely discontinuous Markoff processes, *Trans. Amer. Math. Soc.* **48** (1940), 488–515; Errata **58** (1945), 474.
21. G. E. H. Reuter and W. Lederman, On the differential equations for the transition

probabilities of Markov processes with enumerably many states, *Proc. Cambridge Philos. Soc.* **49** (1953), 247–262.
22. M. JACOBSEN, A characterization of minimal Markov jump processes, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **23** (1972), 32–46.
23. V. A. PERTEL, The greatest lower bound of the variance of an estimate of the instant of disruption of a Poisson flow, *Engrg. Cybernetics* **9** (1971), 926–929.